DATA CENTER OF TOMORROW

Michael Elliott

812 TSS/ENTI

CROSS PLATFORM DATA CENTER

- Background
- Goals
- Milestones
- Foreseen Barriers

CROSS PLATFORM DATA CENTER

“Test & Evaluation data is currently isolated & compartmentalized with little discovery or reuse outside of each respective acquisition program, minimizing its long-term applicability and effectiveness.” TRMC

Cross Platform Data Center is an Air Force Investment Program that has been funded to help address the issue stated above by implementing some of the tools developed by Test Resource Management Center.

WHY CROSS PLATFORM DATA CENTER

- **Most investment has been on *test* needs, but this is an *evaluation* need.**
  - Need an evaluation strategy to not only analyze today's questions but future questions, even after fielding.
- **T&E quality is inadequate for our needs**
  - More data is being collected than can be properly analyzed
  - Only a tiny fraction of data is looked at
  - Only simplistic analysis is being done on a small fraction of data
  - Global view of the collected data is rarely, if ever done
  - No systematic anomaly detection, trend analysis, regression analysis, causality analysis, pattern recognition, simulation/test comparisons, perceived truth/ground truth comparisons are being done.

WHY CROSS PLATFORM DATA CENTER

- **T&E timeliness is inadequate for our needs**
  - Analyst retrieval of test data in some cases takes more than a week
  - Sometimes it's easier (though not cheaper) to just re-run a test rather than find old data that may answer your question
  - Long data ingest times prevent proper debriefing of test participants after a test is over, since their statements cannot be correlated with data in real time.
- **How can T&E be accomplished with improved effectiveness?**
  - Can fewer test be run with an efficient process to analyze data?
  - Cross-program lessons learned are not being made, except anecdotally, can this be changed?

CROSS PLATFORM DATA CENTER

ANYTHING ALARMING ?

Do you have a central location for all types of data related to a single test?

CROSS PLATFORM DATA CENTER

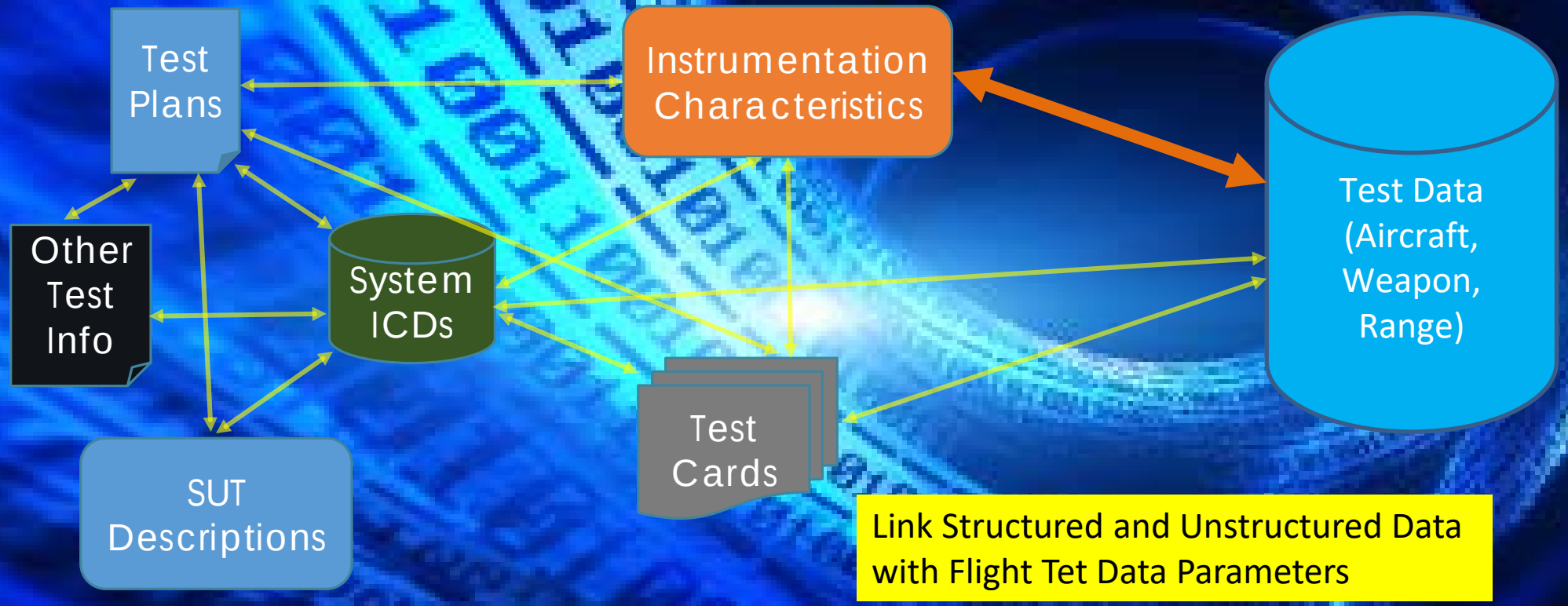
What is Big Data?

Big Data is what ever amount of data that breaks your current system

GOALS OF CPDC- ITEMS WE ARE LOOKING TO ANSWER

- Anomaly Detection – Did something go wrong?
- Causality Detection – What contributed to it?
- Trend Analysis – What's happening over time?
- Predicting Equipment Function and Failure – When will something go wrong?
- Regression Analysis – How is today's data different than the past?
- Data Set Comparison – Are these two large data sets equivalent?
- Pattern Recognition – Are there any recognizable patterns in the data?

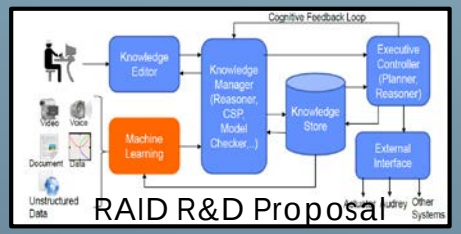
WHAT IS CROSS PLATFORM DATA CENTER



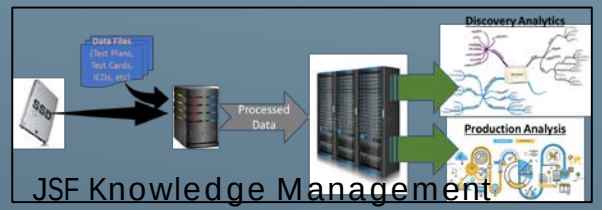
- **Main Objective**
  - **Provide and Sustain a government owned and operated data center with capabilities for the CTFs that will enable data production, data ingest, and data analytics**

CPDC

AUDREY R&D PROJECT



RAID R&D Proposal

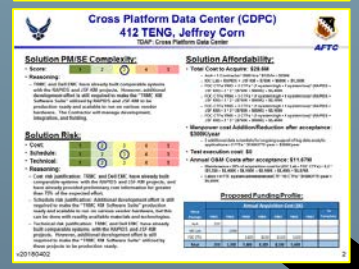


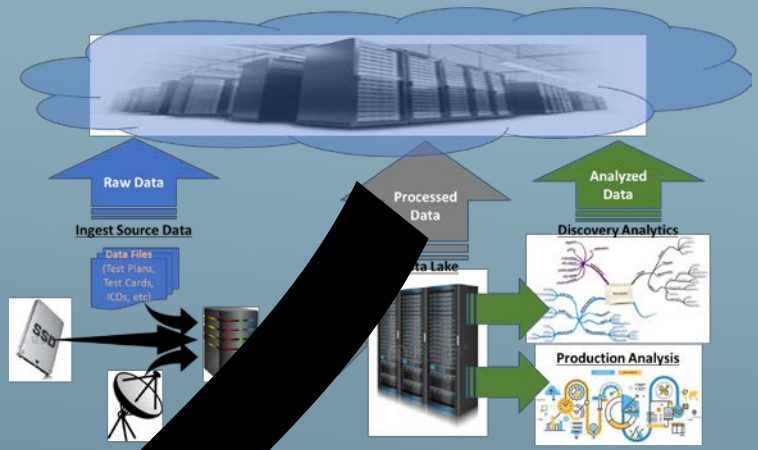
JSF Knowledge Management

Engineering Data Center – Lab Environment

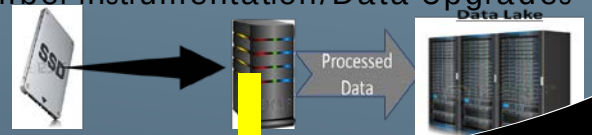
Orange Flag SAPF Lab

Proposed AFTC Investment





Bomber Instrumentation/Data Upgrades



Current

- **Demonstrated Value**
- **Defined Use Cases**
- **Data Ownership Clarification**

- JSF Knowledge Management technology prototyping
- S&T investments in Analytic toolsets
- Development of Experimental Analytic environments
- Increasing Python and Statistics capabilities

- Documenting Requirements
- Developing S&T capabilities
- Leveraging Existing Data Centers

- Implementing AF and DoD Investments in KM
- Continued Tool development

- **AF Investment**
- **Defined Security Plans**
- **Analytics Toolsets Available**

Cross Platform Data Center

Value Package	Annual Acquistion Cost ($M)						
	FY 19	FY20	FY21	FY22	FY23	FY24	To Complete
Requirements	0.5						
IOC Lab	1.18	0.9					
FOC CTF			2.7	5.75	3.65	1.26	
Total	1.68	0.9	2.7	5.75	3.65	1.26	15.94

CPDC PROPOSED PHASES & MILESTONES

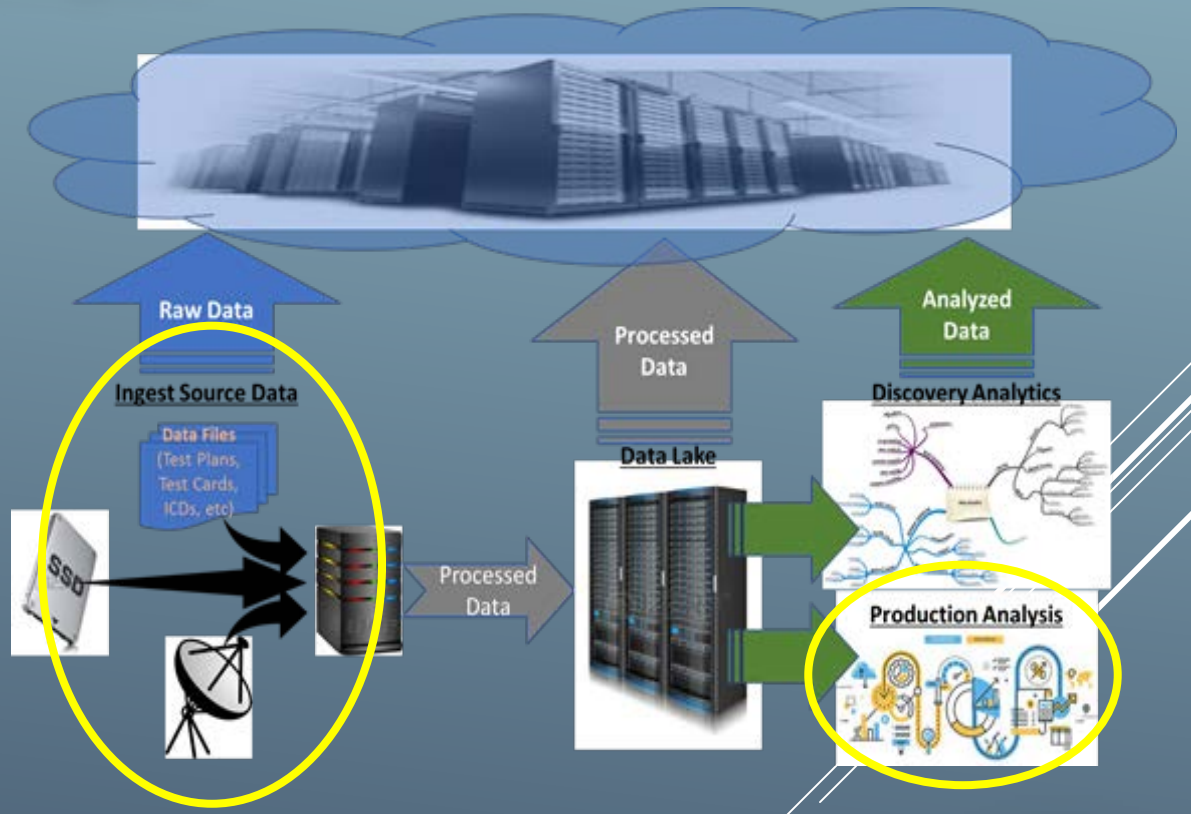
- **Phase 0** – Requirements Gathering
  - Gather customer requirements
  - Implement TRMC Knowledge Management onto Engineering Data Center
- **Phase 1** – Data Production
  - Prototype a software solution on commodity hardware
    - Ingest data into data center
  - Produce data products equivalent to existing data center products
    - Enable Python code to run in prototype environment
- **Phase 2** – Knowledge Management
  - Develop a production system (single classification/caveat)
    - Define system specifications to meet given performance requirements
  - Incorporate ingest of information about test (Plans, Cards, etc.)
    - Link information and test data
  - Develop longitudinal data tools

CPDC PROPOSED PHASES & MILESTONES

- **Phase 3 – Data Analytics**
  - Deploy and configure existing data analytics tools
  - Incorporate Multiple Level Security (Tier 3)
  - Sustain software solution
  - Provide contracts for hardware components

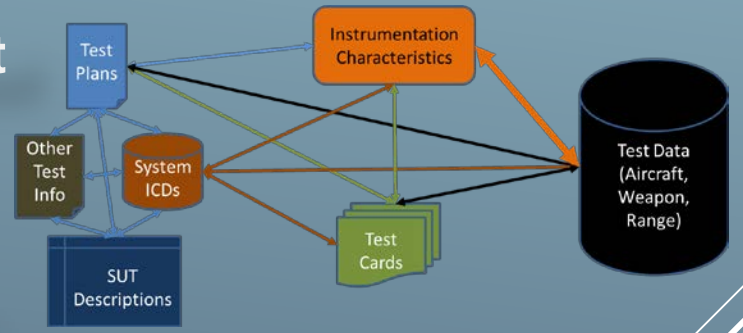
PHASE 1 – DATA PRODUCTION

- 1a – Ingest Data
  - Need to be able to ingest the data from all sources
    - Aircraft
    - Range Sensors
    - Test Project Docs
- 1b – Generate Data Parameters
  - Engineering unit conversion
  - Other 2nd generation data
- 1c – Produce Predefined Reports
  - Already existing reports that are being produced



PHASE 2 – KNOWLEDGE MANAGEMENT

- 2a – Develop "Data Lake" where all of the data is in one spot
  - Test Data and Information about the Test



- 2b – Develop Tools to support "Longitudinal Data Analyses"
  - Example Use Cases
    - Able to compare data from current and previous tests to support regression testing and analyses
    - Able to identify from the historical data all times when an event occurred
      - e.g., Determine all events where F-22 exceeding 3.5 g's
    - Able to augment open air range test data with data from other sources
      - e.g., other Range Sensors, other SUTs, Simulation Runs, HITL Results

PHASE 3 – DATA ANALYTICS

- 3a – Develop and/or Field Data Analytics Tools
  - DoD and Industry has a number of “Analytics” tools that are already available
    - Machine Learning, Neural Networks, Correlation Statistics
    - AUDREY, Python, RAID, JSF-KM
  - Find “Known-Unknowns” and “Unknown-Unknowns”
- 3b – Implement Multi-Level Security
  - Access for appropriate users to data at multiple-caveats

CROSS PLATFORM DATA CENTER

- Foreseen Barriers
  - Security- Multi-Level Security
  - Approval To Operate (ATO)
  - Culture Change
  - Contracts
  - Bureaucracy

CROSS PLATFORM DATA CENTER

Questions?

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
08/05/2019	Briefing slides	May 14-17 2019

4. TITLE AND SUBTITLE
Data Center of Tomorrow
Cross Platform Data Center

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Michael Elliott
Test Resource Management Center

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)
Michael Elliott
307 E. Popson
Edwards AFB, CA 93524

8. PERFORMING ORGANIZATION REPORT NUMBER
412TW-PA-19248

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
412th Test Wing
195 E Popson Ave
Edwards AFB CA 93524

10. SPONSOR/MONITOR'S ACRONYM(S)
N/A

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release A: distribution is unlimited.

13. SUPPLEMENTARY NOTES
International Test & Evaluation Assoc. Workshop, Las Vegas Nevada, May 14-17 2019

14. ABSTRACT
The adoption of Knowledge Management and cloud computing has the potential to be a game changer for the test community and the T&E community has been lagging behind industry in this area. The envisioned solution for this program will provide capabilities for our testers to take advantage of all of the data that they collect by:

- Fast ingest of all the source data for a test program
- Process the data and make it available in a secure data lake environment
- Perform Production Analysis and Discovery Analytics
- Archive Data to a Secure Cloud for long-term storage

Every program will have its challenges, the initial barriers for implementation foreseen thus far are security concerns, quick response purchasing, contracts and a culture shift. The challenges are not trivial, but the solutions are brewing. The data centers of tomorrow will be efficient and effective, delivering the right products when needed.

15. SUBJECT TERMS
Knowledge Management, Cross Platform Data Center, Data Analysis, Big Data Analytics

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					412 TENG/EN (Tech Pubs)
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	None	17	**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified			661-277-8615